Estimating Vaccine Coverage by Using Computer Algebra

Doris Altmann Klaus Altmann

Abstract

The approach of N. Gay for estimating the coverage of a multivalent vaccine from antibody prevalence data in certain age cohorts is improved by using computer aided elimination theory of variables. Hereby, Gay's usage of numerical approximation can be replaced by exact formulas which are surprisingly nice, too.

1 Introduction

(1.1) Nigel Gay [Ga] has estimated the coverage of MMR (measles, mumps, rubella) multivalent vaccination in a fixed age cohort by the following method:
The rates $p(\pm,\pm,\pm)$ of being seropositive with each of the three diseases depend, via a polynomial system F, on the MMR coverage v, the exposition factors e_i, and the rates s_i of seroconversion; the index $i = 1, 2, 3$ stands for measles, mumps and rubella, respectively. On the other hand, it is the $p(\pm,\pm,\pm)$ which can be obtained from the available data. Hence, a maximum likelihood approach provides estimations of v, e_i, and s_i.

Gay's approach leads to numerical methods of finding values v, e_i, s_i that minimize the distance between $F_{\pm,\pm,\pm}(v, e_i, s_i)$ and the measured $p(\pm,\pm,\pm)$. The present paper replaces this part by providing *exact* formulas describing the inverse of the polynomial map $F : \mathbb{R}^7 \to \mathbb{R}^8$. Note that the image of F is contained in the hyperplane $[\sum p(\pm,\pm,\pm) = 1]$, i.e. it is 7-dimensional like the source space of F.
The final result providing our estimation of v, e_i, s_i may be found in Theorem (3.2).

(1.2) We make the same three assumptions used by Gay [Ga]:

(1) Vaccinated children who do not seroconvert as a result of vaccination have the same probability of being seropositive as an unvaccinated child of the same age (i.e., e_i).

(2) In a single individual, seroconversion to each vaccine component is independent.

(3) Risk of exposure to infection is homogeneous within each age cohort and infection with each disease is independent.

However, we eliminate another assumption which is silently made in [Ga] in that we do not assume that the seroconversion s_i for the i-th disease is independent of age.

(1.3) We would like to thank Duco van Straten for the useful discussions concerning the exciting mathematical pattern hidden in the MMR problem and its solution. Moreover, we are greatful to Nigel Gay for sending us his manuscript including the data of the ESEN (European Seroepidemiological Network) Project.

2 The MMR system

(2.1) First, let us recall from [Ga] the involved variables and their mutual relationship. Fixing one of the age cohorts, we denote by

- v the proportion of children who have received the multivalent vaccine ("MMR coverage"),
- e_i the rate measuring the exposure to natural infection with disease i ("exposition factor"),
- s_i the proportion of children previously with no detectable antibody to disease i who acquire detectable antibody to disease i when vaccinated ("seroconversion").

The rate q_i measuring the presence of antibodies to disease i under the condition of being vaccinated may be easily expressed as

$$q_i = e_i + (1 - e_i)\, s_i \quad \text{with} \;\; i = 1, 2, 3.$$

From these data it is possible to obtain information about the expected antibody prevalence in general. It is encoded in the 8 variables $p(\pm, \pm, \pm)$ with "+" at the i-th place standing for the presence and "$-$" for the absence of antibodies to the i-th disease. Likewise, we may think about the sign triples as numbers between 0 (meaning "$- - -$") and 7 (meaning "$+ + +$"); this allows the shorter description $p(\pm, \pm, \pm) = p(k) = p_k$. The equations are

$$\begin{aligned}
p_7 &= p(+,+,+) &&= v\, q_1\, q_2\, q_3 &&+ (1-v)\, e_1\, e_2\, e_3 \\
p_6 &= p(+,+,-) &&= v\, q_1\, q_2\, (1-q_3) &&+ (1-v)\, e_1\, e_2\, (1-e_3) \\
p_5 &= p(+,-,+) &&= v\, q_1\, (1-q_2)\, q_3 &&+ (1-v)\, e_1\, (1-e_2)\, e_3 \\
p_4 &= p(+,-,-) &&= v\, q_1\, (1-q_2)\, (1-q_3) &&+ (1-v)\, e_1\, (1-e_2)\, (1-e_3) \\
p_3 &= p(-,+,+) &&= v\, (1-q_1)\, q_2\, q_3 &&+ (1-v)\, (1-e_1)\, e_2\, e_3 \\
p_2 &= p(-,+,-) &&= v\, (1-q_1)\, q_2\, (1-q_3) &&+ (1-v)\, (1-e_1)\, e_2\, (1-e_3) \\
p_1 &= p(-,-,+) &&= v\, (1-q_1)\, (1-q_2)\, q_3 &&+ (1-v)\, (1-e_1)\, (1-e_2)\, e_3 \\
p_0 &= p(-,-,-) &&= v\, (1-q_1)\, (1-q_2)\, (1-q_3) &&+ (1-v)\, (1-e_1)\, (1-e_2)\, (1-e_3)\,.
\end{aligned}$$

Remark: In [Ga], the variables v, e_i, q_i, and p_k carry a second index pointing to the special age cohort; s_i does not because of Gay's assumption mentioned at the end of (1.2).

(2.2) The previous equations express the variables p_k in terms of v, e_i, q_i or, since $s_i = (q_i - e_i)/(1 - e_i)$, in terms of v, e_i, s_i. Our goal is to describe the inverse dependencies, and we proceed in two steps:
First, using elimination theory, we produce in (2.3) and (2.4) for each of the variables v, e_i, q_i a separate equation with coefficients in the polynomial ring $\mathbb{Q}[p_0, \ldots, p_7]$. The surprising fact will be that all these equations are quadratic ones. Then, as a second step, we will check in (2.5) which of the 2^7 combinations actually provide a solution to our system. The results of these investigations are gathered in Theorem (2.5).

Before we start this program, we would like to introduce an easy technical trick in which we replace the variables p_k by symbolic fractions a_k/n. By doing so, it changes the above equations in the obvious way. For instance, the first one becomes

$$a_7 \;=\; a(+,+,+) \;=\; n\, v\, q_1\, q_2\, q_3 \;+\; n\, (1-v)\, e_1\, e_2\, e_3\,.$$

Since this manipulation increases both the degree and the number of variables, it seemingly complicates the problem. However, using computer algebra systems, the computational time decreases

substantially. Moreover, another advantage of our approach is that $\sum_{k=0}^{7} p_k = 1$ translates into $\sum_{k=0}^{7} a_k = n$. In particular, when finally applying our formulas, we may directly substitute the number of observed probands in each category for the corresponding variables a_k. The number n equals the size of the cohort.

(2.3) Let us start with eliminating n, e_i, q_i to obtain an equation for the variable v which is, by the way, of major interest. We work with the computer algebra system SINGULAR developed at the University Kaiserslautern, [GPS].

Let R be a polynomial ring of characteristic zero with 16 variables a_k, n, v, e_i, q_i. For the monomial order we have to choose a global one, e.g. `dp(16)`. Transforming the 8 equations into an ideal $I \subseteq R$, the command "`eliminate(I,n*e(1)*e(2)*e(3)*q(1)*q(2)*q(3))`" produces a quadratic equation

$$c_1(a_0,\ldots,a_7)\,v^2 - c_1(a_0,\ldots,a_7)\,v + c_0(a_0,\ldots,a_7) \;=\; 0$$

with huge polynomials c_1, c_0 of degree 6 in the variables $a_0,\ldots,a_7$.
We may also use SINGULAR for the factorization of polynomials. Applied to the coefficient c_1 as well as to the discriminant of our quadratic polynomial, this yields nice results. With

$$\begin{aligned}
f_1 \;:=\; n \;&=\; \Big(\big(a_0+a_3+a_5+a_6\big)+\big(a_7+a_4+a_2+a_1\big)\Big)\\
f_3 \;:=\; &\Big(\big(a_0+a_3+a_5+a_6\big)-\big(a_7+a_4+a_2+a_1\big)\Big)\Big(a_0a_7+a_3a_4+a_5a_2+a_6a_1\Big)\\
&-2\Big(a_0a_7(a_0-a_7)+a_3a_4(a_3-a_4)+a_5a_2(a_5-a_2)+a_6a_1(a_6-a_1)\Big)\\
&+2\Big(\big(a_3a_5a_6+a_0a_5a_6+a_0a_3a_6+a_0a_3a_5\big)-\big(a_4a_2a_1+a_7a_2a_1+a_7a_4a_1+a_7a_4a_2\big)\Big)\\
f_4 \;:=\; &\Big(a_0^2a_7^2+a_3^2a_4^2+a_5^2a_2^2+a_6^2a_1^2\Big)\;+\;4\Big(a_0a_3a_5a_6\;+\;a_7a_4a_2a_1\Big)\\
&-2\Big(a_0a_7\,a_3a_4\;+\;a_0a_7\,a_5a_2\;+\;a_0a_7\,a_6a_1\;+\;a_3a_4\,a_5a_2\;+\;a_3a_4\,a_6a_1\;+\;a_5a_2\,a_6a_1\Big),
\end{aligned}$$

we obtain

$$c_1 = f_1^2\,f_4 \qquad \text{and} \qquad c_1 - 4c_0 = f_3^2\,.$$

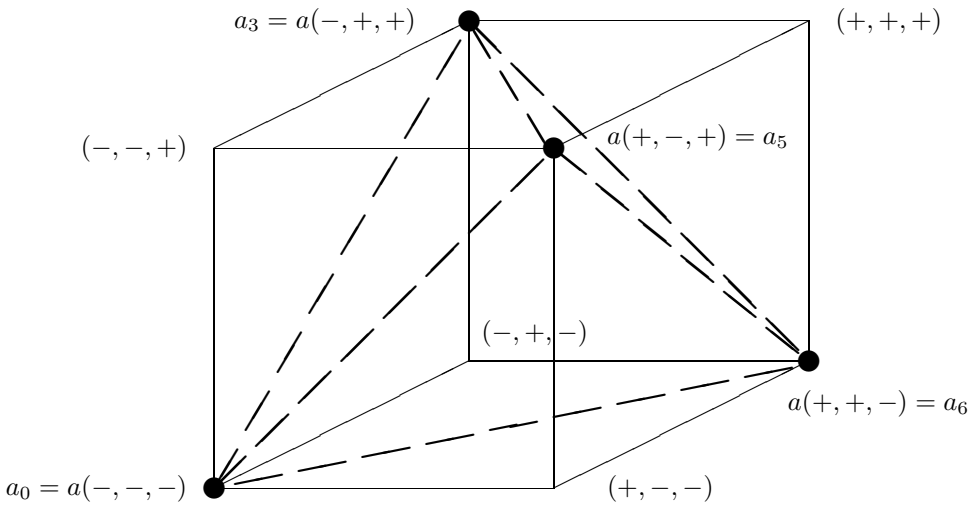
In particular, the two solutions for v are

$$v_{1,2} \;=\; \frac{1}{2}\left(1 \pm \sqrt{\frac{c_1-4c_0}{c_1}}\right) \;=\; \frac{1}{2}\left(1 \pm \frac{f_3(a_0,\ldots,a_7)}{f_1(a_0,\ldots,a_7)\sqrt{f_4(a_0,\ldots,a_7)}}\right).$$

Remarks:

(1) Note that whenever v solves the equation, then so does $(1-v)$. This symmetry may easily be seen in the original 8 equations by switching the variables e_i and q_i.

(2) The formulas for f_1, f_3, and f_4 become very natural if we recall that a_0, a_3, a_5, a_6 correspond to $a(-,-,-)$, $a(-,+,+)$, $a(+,-,+)$, $a(+,+,-)$, respectively. These variables are those which have an even number of plus signs.
This fact may be illustrated by imaging the variables $a(\pm,\pm,\pm)$ as sitting in the corners of a cube. Then, a_0, a_3, a_5, a_6 correspond to the vertices of one of the two inscribed regular tetrahedra. The remaining a_7, a_4, a_2, a_1 are contained in the opposite corners, respectively.



(3) It has been observed by Duco van Straten that f_4 equals the hyperdeterminant of the three-dimensional matrix $A_{\bullet\bullet\bullet}$ formed by the variables $a(\pm,\pm,\pm)$, cf. Proposition 14.1.7. in [GKZ]. Moreover, f_3 is a linear combination of the derivatives of f_4 which follows the usual pattern,

$$2f_3 = \Big(\frac{\partial f_4}{\partial a_0} + \frac{\partial f_4}{\partial a_3} + \frac{\partial f_4}{\partial a_5} + \frac{\partial f_4}{\partial a_6}\Big) - \Big(\frac{\partial f_4}{\partial a_7} + \frac{\partial f_4}{\partial a_4} + \frac{\partial f_4}{\partial a_2} + \frac{\partial f_4}{\partial a_1}\Big)\,.$$

Finally, we would like to note that the coefficient c_0 itself does split into a product of three quadrics:

$$\begin{array}{lllll} c_0 = f_{21}\, f_{22}\, f_{23} & \text{with} & f_{21} & := & (a_0+a_4)(a_7+a_3) - (a_1+a_5)(a_6+a_2) \\ & & f_{22} & := & (a_0+a_2)(a_7+a_5) - (a_4+a_6)(a_3+a_1) \\ & & f_{23} & := & (a_0+a_1)(a_7+a_6) - (a_2+a_3)(a_5+a_4)\,. \end{array}$$

(2.4) Now, we focus on the remaining six variables e_i and s_i. Following the above recipe, we obtain again quadratic equations for each of them, but with much smaller coefficients. They are no longer of degree 6, but quadratic themselves.

Notation: With $A_{\bullet\bullet\bullet}$ being the three-dimensional matrix formed by the variables $a(\pm,\pm,\pm)$, we derive the following ordinary (2×2) matrices from it:

- $A_+(1) := A_{+\bullet\bullet}$ denotes the layer consisting of the entries $a(+,\bullet,\bullet)$, i.e., the right hand face of the cube depicted above; the remaining (left) one forms the matrix $A_-(1) := A_{-\bullet\bullet}$. Similarly, we may define $A_\pm(2) := A_{\bullet\pm\bullet}$ and $A_\pm(3) := A_{\bullet\bullet\pm}$.
- Considering the sum of the layers, we obtain $A_\Sigma(i) := A_+(i) + A_-(i)$ for $i = 1, 2, 3$.

Using this new terminology, we may recover the quadratic c_0-factors f_{2i} from the end of (2.3) as

$$f_{2i} = \det A_\Sigma(i) \qquad \text{with} \qquad i = 1, 2, 3\,.$$

Fixing a disease index i, the elimination done by SINGULAR tells us that e_i and q_i both obey the same quadratic equation. It is

$$\Big(\det A_\Sigma(i)\Big)\, x^2 - \Big(\det A_\Sigma(i) + \det A_+(i) - \det A_-(i)\Big)\, x + \Big(\det A_+(i)\Big) = 0\,.$$

The discriminant is the hyperdeterminant $\det A = f_4$ again. Hence, the solutions for e_i and q_i are

$$\Big[(e_i)_{1,2} \text{ and } (q_i)_{1,2}\Big] \;=\; \frac{1}{2}\left(1+\frac{\det A_+(i)-\det A_-(i)\pm\sqrt{\det A}}{\det A_\Sigma(i)}\right) \;=\; \frac{1}{2}\left(1+\frac{g_{2i}\pm\sqrt{f_4}}{f_{2i}}\right)$$

with g_{2i} being the quadratic polynomials

$$g_{2i} \;:=\; \det A_+(i)-\det A_-(i) = \begin{cases} -a_0a_3+a_1a_2+a_4a_7-a_5a_6 & (\text{for } i=1)\\ -a_0a_5+a_1a_4+a_2a_7-a_3a_6 & (\text{for } i=2)\\ -a_0a_6+a_1a_7+a_2a_4-a_3a_5 & (\text{for } i=3)\,. \end{cases}$$

(2.5) Assuming the general case of $f_1\neq 0$, $\det A_{\bullet\bullet\bullet}\neq 0$, and $\det A_\Sigma(i)\neq 0$ for each $i=1,2,3$, we have narrowed the number of possible values for each of the variables v, e_i, and q_i down to two. It remains to check which of the 2^7 combinations survive to provide an actual solution of the original system (2.1).
This can easily be done by considering the sum of those equations out of the original system that correspond to a certain face of the cube depicted in (2.3). For instance, adding up the equations for a_7, a_6, a_5, and a_4 provides

$$a_7+a_6+a_5+a_4 \;=\; f_1\,v\,q_1 \;+\; f_1\,(1-v)\,e_1\,.$$

All variables have been eliminated except v, q_1, and e_1. This allows us to show that the e's must not equal the q's. (Assuming $e_1=q_1$, we would obtain $a_7+a_6+a_5+a_4 = f_1e_1$. However, substituting this value of e_1 into the quadratic equation of (2.4) yields

$$f_1^2\Big(f_{21}\,{e_1}^2-(f_{21}+g_{21})\,e_1+\det A_+(1)\Big) \;=\; -\,f_{22}\,f_{23}\,,$$

which is generally different from zero.)
Now, by Remark (2.3)(1), we may assume that, w.l.o.g., $v=(f_1\sqrt{f_4}+f_3)/(2f_1\sqrt{f_4})$. Hence, with $e_1=(f_{21}+g_{21}\mp\sqrt{f_4})/(2f_{21})$ and $q_1=(f_{21}+g_{21}\pm\sqrt{f_4})/(2f_{21})$, the above equation multiplied with $4f_{21}\sqrt{f_4}$ becomes

$$\begin{aligned} 4\,f_{21}\,\sqrt{f_4}\Big(\sum_{k=4}^{7}a_k\Big) \;&=\; \Big(f_1\sqrt{f_4}+f_3\Big)\Big(f_{21}+g_{21}\pm\sqrt{f_4}\Big)+\Big(f_1\sqrt{f_4}-f_3\Big)\Big(f_{21}+g_{21}\mp\sqrt{f_4}\Big)\\ &=\; 2\,f_1\,\sqrt{f_4}\,\big(f_{21}+g_{21}\big)\;\pm\;2\,f_3\,\sqrt{f_4}\,. \end{aligned}$$

In particular, since $2f_{21}\big(\sum_{k=4}^{7}a_k\big) = f_1\big(f_{21}+g_{21}\big)+f_3$, only the signs on top survive in the formulas of e_1 and q_1.
Finally, one may use SINGULAR again for checking that these values, together with the similar ones for the remaining variables, indeed yield a solution of the original system. This means that we have shown the following

Theorem: *If $f_1, f_4, f_{2i}\neq 0$ for $i=1,2,3$, then the polynomial system of (2.1), with the adaption $p_k=a_k/n$ made in (2.2), has exactly two solutions. They are*

$$v=\frac{f_1\,\sqrt{f_4}\;\pm\;f_3}{2\,f_1\,\sqrt{f_4}}\,,\quad e_i=\frac{f_{2i}+g_{2i}\mp\sqrt{f_4}}{2\,f_{2i}}\,,\quad q_i=\frac{f_{2i}+g_{2i}\pm\sqrt{f_4}}{2\,f_{2i}}\qquad (i=1,2,3).$$

If some of the above polynomials $f_\bullet$ do vanish, then the system (2.1) might have infinitly many solutions or no solution at all.

3 The MMR coverage

(3.1) If we apply the previous theory to our statistical problem of estimating the MMR coverage, then a_k stands for the number of persons of a prefixed age group observed to have antibody status k ($k = 0, \ldots, 7$). Thus, f_1 is the size of the cohort, and this number is automatically positive. On the other hand, we would like to interpret the solutions v, e_i, q_i, and s_i of the MMR system as estimations of the probabilities described in (2.1). In particular, they should be real numbers and, moreover, be contained in the interval $[0, 1]$.
While in [Ga] the latter is forced by the numerical program used to solve the system, our solutions may not have these properties. However, this should not be considered problematic, but a feature of our method. If the solutions fall out of the range making sense, this is a strong hint that the input data a_k are of poor quality.

(3.2) In the following, we will formulate the conditions the input data have to fulfill for yielding apropriate results. Moreover, we will see that, in the statistical context, only one of the two solutions mentioned in Theorem (2.5) survives.

Theorem: *Let a_k be the observed number of people in a fixed age group with antibody status k. Then, the MMR system has a good statistical solution if and only if*

$$f_4(\underline{a}) > 0 \quad \textit{and} \quad f_{2i}(\underline{a}) \geq \sqrt{f_4(\underline{a})} + \left|g_{2i}(\underline{a})\right| \quad (i = 1, 2, 3)\,.$$

If these conditions are satisfied, then the estimation for v, e_i, s_i is

$$v = \frac{f_1\sqrt{f_4} + f_3}{2\,f_1\sqrt{f_4}}\,, \quad e_i = \frac{f_{2i} + g_{2i} - \sqrt{f_4}}{2\,f_{2i}}\,, \quad s_i = \frac{2\sqrt{f_4}}{f_{2i} - g_{2i} + \sqrt{f_4}} \qquad (i = 1, 2, 3).$$

Proof: Positivity of f_4 means that the solutions described in Theorem (2.5) are real. Assuming this, we have

$$v \in [0, 1] \iff f_1\sqrt{f_4} \pm f_3 \geq 0 \iff f_1^2 f_4 \geq f_3^2\,.$$

On the other hand, we have seen in (2.3) that

$$f_1^2 f_4 \;=\; c_1 \;=\; (c_1 - 4c_0) + 4c_0 \;=\; f_3^2 \;+\; 4\,f_{21}\,f_{22}\,f_{23}\,.$$

Hence, the condition "$v \in [0, 1]$" is equivalent to $f_{21}f_{22}f_{23} > 0$.
Since $s_i = (q_i - e_i)/(1 - e_i)$, we know that

$$e_i,\, s_i \in [0, 1] \iff 0 \leq e_i \leq q_i \leq 1\,.$$

From Theorem (2.5) we obtain, depending on the choice of the solution, that $q_i - e_i = \sqrt{f_4}/f_{2i}$ for $i = 1, 2, 3$ or that $q_i - e_i = -\sqrt{f_4}/f_{2i}$ for $i = 1, 2, 3$. Anyway, for $q_i \geq e_i$, the polynomials f_{21}, f_{22}, f_{23} must have the same sign. Together with $f_{21}f_{22}f_{23} > 0$ obtained above, this means that $f_{21}, f_{22}, f_{23} > 0$. In particular, looking at Theorem (2.5), only the solution with the top sign survives.
Finally, it is easy to see that the conditions $e_i \geq 0$ and $q_i \leq 1$ translate into $f_{2i} \geq \sqrt{f_4} - g_{2i}$ and $f_{2i} \geq \sqrt{f_4} + g_{2i}$, respectively. □

(3.3) **Remark:** If one is only interested in the MMR coverage v, then the conditions ensuring a meaningful result may be weakened. It follows from the proof of the previous theorem that

$$f_4(\underline{a}) > 0 \quad \text{and} \quad f_{2i}(\underline{a}) > 0 \quad (i = 1, 2, 3)\,.$$

will do.

4 Data

(4.1) To illustrate our results, we have chosen some data of some country of the ESEN Project, [Ga]. These data have not yet been finalized as they might be changed according to a new standardization between the European countries. For that reason, the use of these data here is for illustrative purposes only.
The input, i.e., the sampled variables a_k, may be found in the table (4.2). The first table compares our estimation of v, e_1, e_2, and e_3 by age groups (AG) with that obtained by Gay in [Ga]; the variables pointing to his values carry a tilde.

AG	$\tilde{v}$	v	$\tilde{e}_1$	e_1	$\tilde{e}_2$	e_2	$\tilde{e}_3$	e_3	s_1	s_2	s_3
1	0.227	0.227	0.003	0.005	0.019	0.019	0.014	0.011	0.950	0.861	0.974
2	0.642	0.642	0.122	0.144	0.020	0.017	0.090	0.090	0.976	0.878	0.922
3	0.715	0.710	0.122	0.112	0.041	0.046	0.090	0.087	1.002	0.912	0.930
4	0.837	0.824	0.251	0.279	0.041	0.054	0.106	0.219	1.003	0.886	0.922
5	0.859	0.863	0.292	0.252	0.241	0.227	0.106	0.000	1.000	0.886	0.921
6	0.794	0.889	0.621	0.427	0.324	0.094	0.106	-0.037	0.961	0.855	0.830
7	0.645	0.847	0.756	0.550	0.502	0.006	0.256	0.258	0.949	0.938	0.678
8	0.662	0.794	0.764	0.652	0.502	0.285	0.411	0.356	0.969	0.877	0.798
9	0.576	0.900	0.764	0.588	0.665	0.279	0.481	-0.007	0.833	0.857	0.838
10	0.478	0.940	0.906	0.667	0.734	0.049	0.631	0.450	0.906	0.892	0.660

The main difference between Gay's and our results can be found in the values of v, e_1, e_2, e_3 in the higher age groups.
Moreover, while Gay has assumed age independent seroconversion rates, our solutions s_1, s_2, s_3 do vary with age; the most striking example is the rubella seroconversion s_3. The comparison of Gay's values with the age average of our solutions for s_1, s_2, s_3 is as follows:

Seroconversion by N. Gay:	0.989	0.880	0.910
Average of our s_1, s_2, s_3:	0.955	0.884	0.847

(4.2) We can use the equations of (2.1) to re-calculate the expected antibody prevalence out of the solutions obtained for v, e_i, s_i. In other words, for each antibody status $(\pm, \pm, \pm)$ we are looking for the number of people that should have been observed to yield the desired result.
Because we used an exact method, it is no surprise that our solutions give exactly back the input data; they fill the a_k-columns in the following table. On the other hand, using Gay's solutions, we obtain different values which are contained in the $\tilde{a}_k$-columns:

$---$		$--+$		$-+-$		$-++$		$+--$		$+-+$		$++-$		$+++$	
$\tilde{a}_0$	a_0	$\tilde{a}_1$	a_1	$\tilde{a}_2$	a_2	$\tilde{a}_3$	a_3	$\tilde{a}_4$	a_4	$\tilde{a}_5$	a_5	$\tilde{a}_6$	a_6	$\tilde{a}_7$	a_7
155.8	156	2.3	2	3.1	3	0.5	2	1.0	1	5.0	6	3.7	1	37.7	38
49.1	48	5.0	5	1.1	1	1.0	2	7.9	9	12.7	13	8.2	7	90.2	90
40.8	42	4.2	4	1.8	2	1.2	0	6.9	6	14.6	11	9.8	8	107.6	114
20.1	18	2.5	5	1.0	1	1.2	0	8.2	8	17.7	18	11.6	9	129.7	133
14.6	17	1.8	0	4.7	5	1.8	0	7.3	7	16.1	15	15.3	15	153.4	156
10.2	13	1.3	0	5.0	2	1.2	3	17.9	14	14.8	20	20.7	30	145.9	135
6.9	11	2.4	4	7.0	1	2.7	3	21.9	16	15.1	13	30.3	40	128.7	127
5.0	7	3.5	4	5.0	3	3.8	3	16.5	15	19.1	20	23.2	25	135.9	135
3.4	6	3.1	1	6.7	4	6.5	9	11.1	11	14.4	14	26.7	27	122.1	122
0.9	2	1.5	2	2.4	1	4.2	4	8.5	7	17.1	17	26.1	28	121.2	121

(4.3) In the following, we will discuss some of the properties of our solutions.

(1) One should not so much worry about negative rates or rates above 1 as they appear among the e_i or s_i. In all those cases, the values are very close to the allowed range.

(2) Our major concern is caused by the exposition factors e_2 and e_3. They seem to be very small in the higher age groups and, additionally, they do not increase with age.
For the latter, however, we may use the same explanation as Gay did for the decline of his v in older cohorts in that the data arise from *different* cohorts in each age group.

(3) As already mentioned before, we did not ad hoc assume that the seroconversions s_i are age independent. However, as a result of our calculations, we obtained values for mumps and measles that did not greatly vary – and the averages are quite close to Gay's values.
On the other hand, the seroconversion factor for rubella shows an unusual behavior in the higher age groups and we would be interested in an explanation for it.

The major difference between Gay's and our approach is the following:
Altmann: We consider each age group separately; this yields a system of 7 equations in 7 variables for each group, allowing exact solutions with easy formulas.
Gay: He considers 10 age groups at once, yielding a system with 70 equations in 70 variables. Moreover, he creates additional restrictions by

- assuming that the seroconversion s_i is age independent (meaning to lose 27 variables),
- and by forcing the exposition factors e_i to increase with age (meaning to introduce additional inequalities).

For the remaining system, Gay uses a numerical approach to find values for v(age), e_i(age), and s_i to fit into the system as best as possible. Exact solutions are of course out of range.

Thus, the fact that the above problem (2) does not occur in Gay's solutions is no surprise at all. It was part of his method to force all these properties which are, however, biologically plausible. An advantage of Gay's method is that imperfect data in single age groups might be corrected by the better ones.
On the other hand, our method tells which data are better or worse and gives information about their quality. Moreover, besides exactness, the main advantage of our approach seems to be that the formulas for v, e_i, s_i are mutually independent. Hence, even if one dislikes the results for the e_i's or s_i's, one has still an explicit formula for the MMR coverage v which works well.

References

[Ga] Gay, N.: A Method for Estimating Coverage of a Multivalent Vaccine from Antibody Prevalence Data: application to MMR vaccine in 3 European countries. Draft.

[GKZ] Gelfand, I.M., Kapranov, M.M., Zelevinsky, A.V.: Discriminants, Resultants, and Multidimensional Determinants. Birkhäuser Boston 1994

[GPS] Greuel, G.-M., Pfister, G., Schönemann, H.: Singular. System for computer algebra, university of Kaiserslautern, available via `www.mathematik.uni-kl.de`

Doris Altmann
Robert Koch Institut
Stresemannstr. 90-102
D-10963 Berlin, Germany
e-mail: altmannd@rki.de

Klaus Altmann
Institut für Reine Mathematik
Humboldt-Universität zu Berlin
Ziegelstr. 13A
D-10099 Berlin, Germany
e-mail: altmann@mathematik.hu-berlin.de